

Appln No. 10/524,606

Amdt date June 5, 2007

Reply to Office action of February 5, 2007

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A method of classifying an object into one of a number n of classes wherein n is 2 or more, wherein information concerning said object is provided in a plurality of test data T, ~~determining whether a test sample, having test data T, is categorized in one of a number n of classes wherein n is 2 or more, the method~~ comprising the steps of:

extracting a plurality of emerging patterns from a training data set D that has at least one instance of each of ~~[[said]]~~ n classes of data corresponding to said n classes, wherein each of said emerging patterns comprises a plurality of conjunctive conditions, each condition having a variable and a constraint, and wherein a plurality of occurrences satisfy said conditions for an i th one of said n classes of data, but no occurrence satisfies said conditions for a remainder of said n classes of data;

creating n lists, wherein:

an i th list of said n lists contains a frequency of occurrence, $f_i(m)$, of each emerging pattern $EP_i(m)$ from said plurality of emerging patterns that has a non-zero occurrence in an i th class of data;

using a fixed number, k , of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating n scores; wherein:

an i th score corresponding to an i th class of said n scores is derived from the frequencies of k emerging patterns in said i th list that also occur in said test data; and

deducing which of said n classes ~~of data~~ the test data object is categorized in, by selecting the highest of said n scores.

Appln No. 10/524,606

Amdt date June 5, 2007

Reply to Office action of February 5, 2007

2. (Currently Amended) The method of claim 1, additionally comprising:
if there is more than one class with the highest score, deducing which of said n classes of data the ~~test data~~object is categorized in by selecting the largest of the classes of data having the highest score.

3. (Previously Presented) The method of claim 1, wherein:
said k emerging patterns of the i th list that occur in said test data have the highest frequencies of occurrence in said i th list amongst all those emerging patterns of said i th list that occur in said test data, for all i .

4. (Previously Presented) The method of claim 1, wherein:
emerging patterns in the i th list are ordered in descending order of said frequency of occurrence in said i th class of data, for all i .

5. (Previously Presented) The method of claim 1, wherein the i th list has a length l_i , and k is a fixed percentage of the smallest l_i .

6. (Previously Presented) The method of claim 1, wherein the i th list has a length l_i , and k is a fixed percentage of $\sum_{i=1}^n l_i$

7. (Previously Presented) The method of claim 1, wherein the i th list has a length l_i , and k is a fixed percentage of any l_i .

8. (Previously Presented) The method of claim 5, wherein said fixed percentage is from about 1% to about 5% and k is rounded to a nearest integer value.

9. (Previously Presented) The method of claim 1, wherein $n = 2$.

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

10. (Previously Presented) The method of claim 1, wherein $n = 3$ or more.

11. (Currently Amended) A method of ~~determining whether~~ classifying an object into a first class or a second class, wherein information concerning said object is provided in a plurality of a test sample, having test data T, is categorized in a first class or a second class, the method comprising the steps of:

extracting a plurality of emerging patterns from a training data set D that has at least one instance of a first class of data corresponding to said first class and at least one instance of a second class of data corresponding to said second class, wherein each of said emerging patterns comprises a plurality of conjunctive conditions, each condition having a variable and a constraint, and wherein a plurality of occurrences satisfy said conditions for said first class of data, but no occurrence satisfies said conditions for said second class of data, or wherein an alternate plurality of occurrences satisfies said conditions for said second class of data, but no alternate occurrence satisfies said conditions for said first class of data;

creating a first list and a second list wherein:

said first list contains a frequency of occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in said first class of data; and

said second list contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in said second class of data;

using a fixed number, k , of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating:

a first score corresponding to said first class, said first score derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score corresponding to said second class, said second score derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

deducing whether the ~~test data~~object is categorized in the first class of data or in the second class ~~of data~~ by selecting the higher of said first score and said second score.

12. (Currently Amended) The method of claim 11, additionally comprising:
if said first score and said second score are equal, deducing whether the ~~test sample~~object is categorized in the first class ~~of data~~ or in the second class ~~of data~~ by selecting the larger of the first or the second class of data.

13. (Previously Presented) The method of claim 11, wherein:
said k emerging patterns of said first list that occur in said test data have the highest frequencies of occurrence in said first list amongst all those emerging patterns of said first list that occur in said test data; and
said k emerging patterns of said second list that occur in said test data have the highest frequencies of occurrence in said second list amongst all those emerging patterns of said second list that occur in said test data.

14. (Previously Presented) The method of claim 11, wherein:
emerging patterns in said first list are ordered in descending order of said frequency of occurrence in said first class of data, and
emerging patterns in said second list are ordered in descending order of said frequency of occurrence in said second class of data.

15. (Previously Presented) The method of claim 11, additionally comprising:
creating a third list and a fourth list, wherein:
said third list contains a frequency of occurrence, $f_1(i_m)$, in said first class of data of each emerging pattern i_m from said plurality of emerging patterns that has a non-zero occurrence in said first class of data and which also occurs in said test data; and
said fourth list contains a frequency of occurrence, $f_2(j_m)$, in said second class of

Appln No. 10/524,606

Amdt date June 5, 2007

Reply to Office action of February 5, 2007

data of each emerging pattern j_m from said plurality of emerging patterns that has a non-zero occurrence in said second class of data and which also occurs in said test data; and wherein emerging patterns in said third list are ordered in descending order of said frequency of occurrence in said first class of data, and

emerging patterns in said fourth list are ordered in descending order of said frequency of occurrence in said second class of data.

16. (Previously Presented) The method of claim 15, wherein:

said first score is given by: $\sum_{m=1}^k \frac{f_1(i_m)}{f_1(m)} \Big|_{EP_1(i_m) \in T}$; and

said second score is given by: $\sum_{m=1}^k \frac{f_2(j_m)}{f_2(m)} \Big|_{EP_2(j_m) \in T}$.

17. (Previously Presented) The method of claim 11, wherein said first list has a length l_1 , and said second list has a length l_2 , and k is a fixed percentage of whichever of l_1 and l_2 is smaller.

18. (Previously Presented) The method of claim 11, wherein said first list has a length l_1 , and said second list has a length l_2 , and k is a fixed percentage of a sum of l_1 and l_2 .

19. (Previously Presented) The method of claim 11, wherein said first list has a length l_1 , and said second list has a length l_2 , and k is a fixed percentage of any one of l_1 or l_2 .

20. (Previously Presented) The method of claim 17, wherein said fixed percentage is from about 1% to about 5% and k is rounded to a nearest integer value.

21. (Previously Presented) The method of claim 1, wherein k is from about 5 to

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

about 50.

22. (Previously Presented) The method of claim 21, wherein k is about 20.
23. (Previously Presented) The method of claim 1, wherein each emerging pattern is expressed as a conjunction of conditions.
24. (Previously Presented) The method of claim 1, wherein only left boundary emerging patterns are used.
25. (Previously Presented) The method of claim 1, wherein only plateau emerging patterns are used.
26. (Previously Presented) The method of claim 25 wherein only the most specific plateau emerging patterns are used.
27. (Previously Presented) The method of claim 1, wherein each of said emerging patterns has a growth rate larger than a threshold, τ .
28. (Previously Presented) The method of claim 27 wherein said threshold is from about 2 to about 10.
29. (Previously Presented) The method of claim 1, wherein each of said emerging patterns has a growth rate of ∞ .
30. (Previously Presented) The method of claim 1, additionally comprising discretizing said data set, before said extracting.

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

31. (Previously Presented) The method of claim 30, wherein said discretizing utilizes an entropy-based method.

32. (Previously Presented) The method of claim 30, additionally comprising applying a method of correlation based feature selection to said data set, after said discretizing.

33. (Previously Presented) The method of claim 30 additionally comprising applying a chi-squared method to said data set, after said discretizing.

34. (Currently Amended) The method of claim 1, wherein said data set comprises gene expression data, and wherein said object comprises a person.

35. (Previously Presented) The method of claim 34, wherein said gene expression data has been acquired from a micro-array apparatus.

36. (Currently Amended) The method of claim 1, wherein at least one class of data corresponds to data for a first type of cell and at least another class of data corresponds to data for a second type of cell, and wherein said object comprises a person.

37. (Original) The method of claim 36, wherein said first type of cell is a normal cell and said second type of cell is a cancerous cell.

38. (Currently Amended) The method of claim 1, wherein at least one class of data corresponds to data for a first population of subjects and at least another class of data corresponds to data for a second population of subjects, and wherein said object comprises at least one person.

39. (Currently Amended) The method of claim 1, wherein said data set

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

comprises patient medical records, and said object comprises a patient.

40. (Currently Amended) The method of claim 1, wherein said data set comprises financial transactions, and wherein said object comprises at least one person.

41. (Currently Amended) The method of claim 1, wherein said data set comprises census data, and wherein said object comprises at least one person.

42. (Currently Amended) The method of claim 1, wherein said data set comprises characteristics of an item selected from the group consisting of: a foodstuff; an article of manufacture; and a raw material, and wherein said object comprises said item.

43. (Previously Presented) The method of claim 1, wherein said data set comprises environmental data.

44. (Previously Presented) The method of claim 1, wherein said data set comprises meteorological data.

45. (Currently Amended) The method of claim 1, wherein said data set comprises characteristics of a population of organisms, and wherein said object comprises at least one of a similar type of said population of organisms.

46. (Previously Presented) The method of claim 1, wherein said data set comprises marketing data.

47-56. (Cancelled)

57. (Currently Amended) A system for ~~determining whether~~ classifying an

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

object into a first class or a second class, wherein information concerning said object is provided in a plurality of a test sample, having test data T, is categorized in a first class or a second class,
the system comprising:

at least one memory, at least one processor and at least one user interface, all of which are connected to one another by at least one bus;

wherein said at least one processor is configured to:

access a data set that has at least one instance of a first class of data corresponding to said first class, and at least one instance of a second class of data corresponding to said second class;

extract a plurality of emerging patterns from said data set, wherein each of said emerging patterns comprises a plurality of conjunctive conditions, each condition having a variable and a constraint, and wherein a plurality of occurrences satisfy said conditions for said first class of data, but no occurrence satisfies said conditions for said second class of data, or wherein an alternate plurality of occurrences satisfies said conditions for said second class of data, but no alternate occurrence satisfies said conditions for said first class of data;

create a first list and a second list wherein, for each of said plurality of emerging patterns:

said first list contains a frequency of occurrence, $f_i^{(1)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said first class of data, and

said second list contains a frequency of occurrence, $f_i^{(2)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said second class of data;

use a fixed number, k , of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, to calculate:

a first score corresponding to said first class, wherein said first score is derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score corresponding to said second class, wherein said second score is

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

deduce whether the ~~test sample~~object is categorized in the first class ~~of data~~ or in the second class ~~of data~~ by selecting the higher of the first score and the second score.

58. (Currently Amended) The system of claim 57, wherein said processor is additionally configured to:

if said first score and said second score are equal, deduce whether the ~~test sample~~object is categorized in the first class ~~of data~~ or in the second class ~~of data~~ by selecting the larger of the first or the second class of data.

59. (Previously Presented) The system of claim 57, wherein:
said k emerging patterns of said first list that occur in said test data have the highest frequencies of occurrence in said first list amongst all those emerging patterns of said first list that occur in said test data; and

said k emerging patterns of said second list that occur in said test data have the highest frequencies of occurrence in said second list amongst all those emerging patterns of said second list that occur in said test data.

60. (Previously Presented) The system of claim 57 wherein said processor is additionally configured to:

order emerging patterns in said first list in descending order of said frequency of occurrence in said first class of data, and

order emerging patterns in said second list in descending order of said frequency of occurrence in said second class of data.

61. (Previously Presented) The system of claim 57, wherein said processor is additionally configured to:

create a third list and a fourth list, wherein:

said third list contains a frequency of occurrence, $f_1(i_m)$, in said first class of data of each emerging pattern i_m from said plurality of emerging patterns that has a non-zero occurrence in said first class of data and which also occurs in said test data; and

said fourth list contains a frequency of occurrence, $f_2(j_m)$, in said second class of data of each emerging pattern j_m from said plurality of emerging patterns that has a non-zero occurrence in said second class of data and which also occurs in said test data; and wherein

emerging patterns in said third list are ordered in descending order of said frequency of occurrence in said first class of data, and

emerging patterns in said fourth list are ordered in descending order of said frequency of occurrence in said second class of data.

62. (Previously Presented) The system of claim 61, wherein said processor is additionally configured to calculate:

said first score according to the formula: $\sum_{m=1}^k \frac{f_1(i_m)}{f_1(m)} \Big|_{EP_1(i_m) \in T}$; and

said second score according to the formula: $\sum_{m=1}^k \frac{f_2(j_m)}{f_2(m)} \Big|_{EP_2(j_m) \in T}$.

63. (Previously Presented) The system of claim 57, wherein k is from about 5 to about 50.

64. (Previously Presented) The system of claim 57, wherein only left boundary emerging patterns are used.

65. (Previously Presented) The system of claim 57, wherein each of said emerging patterns has a growth rate of ∞ .

Appln No. 10/524,606

Amdt date June 5, 2007

Reply to Office action of February 5, 2007

66. (Previously Presented) The system of claim 57, wherein said data set comprises data selected from the group consisting of: gene expression data, patient medical records, financial transactions, census data, characteristics of an article of manufacture, characteristics of a foodstuff, characteristics of a raw material, meteorological data, environmental data, and characteristics of a population of organisms.

67. (Currently Amended) A method of determining whether a sample cell is cancerous, comprising:

extracting a plurality of emerging patterns from a data set that comprises gene expression data for a plurality of cancerous cells and a gene expression data for a plurality of normal cells, wherein each of said emerging patterns comprises a plurality of conjunctive conditions, each condition having a variable and a constraint, and wherein a plurality of occurrences satisfy said conditions for an i th one of said n classes of data, but no occurrence satisfies said conditions for a remainder of said n classes of data;

creating a first list and a second list wherein:

said first list contains a frequency of occurrence, $f_i^{(1)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said cancerous cells, and

said second list contains a frequency of occurrence, $f_i^{(2)}$, of each emerging pattern i from said plurality of emerging patterns that has a non-zero occurrence in said normal cells;

using a fixed number, k , of emerging patterns, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating:

a first score derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and

a second score derived from the frequencies of k emerging patterns in said second

Appln No. 10/524,606
Amdt date June 5, 2007
Reply to Office action of February 5, 2007

list that also occur in said test data; and

deducing whether the sample cell is cancerous if said first score is higher than said second score.

68-75. (Cancelled)